

呈现科研数据知识库: re3data.org 注册机制

Heinz Pampel[1] Paul Vierkant[2] Frank Scholze[3] Roland Bertelmann[1] Maxi Kindling[2] Jens Klump[1] Hans-Jürgen Goebelbecker[3] Jens Gundlach[3] Peter Schirmbacher[2] Uwe Dierolf [3] (著)

[1](Deutsches GeoForschungsZentrum GFZ, Library and Information Services (LIS), Potsdam 14473, Germany)

[2](Humboldt-Universität zu Berlin, Berlin School of Library and Information Science, Berlin 10099, Germany)

[3](Karlsruhe Institute of Technology (KIT), KIT Library, Karlsruhe 76131, Germany)

顾立平(译)

(中国科学院国家科学图书馆 北京 100190)

摘要: 科研人员在钻研科研问题与分享科研数据的过程中, 需要某种基础设施来确保数据最大程度的获取性、稳定性和可用性。这类基础设施可以统称为科研数据知识库(Research Data Repository, RDR)。自 2012 年启动的 re3data.org 项目, 主要从事科研数据知识库的登记注册, 以及为科研人员、科研资助组织、图书馆和出版商等提供有关异构科研数据知识库的全景概述。截至 2013 年 7 月, 已有 400 个科研数据知识库向 re3data.org 登记, 其中 288 个采用 re3data.org 的信息图标, 以协助科研人员遴选合适的知识库, 并且存储与重用他们的数据。这篇论文描绘异构 RDR 的全景, 表述机构的、学科的、跨学科的以及项目专业的 RDR 类型。深入描述 re3data.org 的特性, 以及这套注册系统如何协助科研人员分辨适合存储和搜索科研数据的知识库。

关键词: 科研数据 科研数据管理 开放获取 机构知识库发展 开放科学 科研数据知识库 信息管理

分类号: G250

1 引 言

科研数据共享与重用及其发展, 近年来逐渐得到重视。其实, 早在 2003 年世界主要科研组织就开始呼吁科学与人文知识的开放获取, 而且在其《柏林宣言》(Berlin Declaration, 2003)中, 也将科研数据作为学术知识的一部分[1]。在 2007 年经济合作发展组织(OECD)发布的《公共资金科研数据获取的原则与指南》中, 提出"促进科研人员之间的数据访问和共享"作为宗旨[2]。这是在学术体系中涉及不同利益相关者的广泛持续辩论中, 可被人们遵循的两份早期参考文献。

英国皇家学会在 2012 年发布《科学是开放事业》的报告中, 敦促科学家们在"可理解的开放性(Intelligent Openness)"的层面上, 使得科研人员的科研数据可获取和可利用: "科学家们应该在合适的数据知识库里存储数据, 让人使用和验证数据"[3]。欧盟的一项政策回应了这个建议: 欧盟寻求成员国通过有关政策以确保"公共资金资助科研成果的科研数据通过电子基础设施实现公共获取、使用和重用" [4]。美国政府更进一步, 要求美国国立科研机构的数字科研数据实施最大程度的获取。白宫科技政策办公室(OSTP)指明"无论全部或者部分受到联邦资助的科研项目, 所产生的数字形式的科研数据都应该存储起来, 并且提供搜寻、检索和使用等的公共访问获取" [5]。欧盟计划在第 8 次 Horizon 2020 框架中加入

收稿日期: 2013-11-29

*本文译自: Making Research Data Repositories Visible: The re3data.org Registry,PLOS ONE,2013,8(11):e78080.DOI: 10.1371/journal.pone.007808, 经原作者授权翻译以中文发表。

类似的要求[6]。上述政策的具体执行，有赖于在不同学科中对科研数据知识库进行取舍，而这就要求科研数据术语的准确定义。

根据科研方法和不同学科中的科研对象的特征，人们对科研数据会有不同的认识。然而，科研数据的概念检验，如同科研数据知识库一样重要，必须服务不同学院和学术社群以及符合它们对科研数据概念的认识。信息基础设施的需求同样来自于内容和用户需求。

所以，我们定义科研数据这个术语，即下：作为科研过程的一部分或者结果的数字数据。

此过程覆盖科研的所有阶段，从科研数据生成到科研结果产出，包括科学、社会科学实证研究或是文化现象观察等。数字科研数据因为学科和方法的不同，而有不同的数据类型、不同程度的集成方式以及数据格式。就科研数据使用与重用的获取目的而言，如果缺乏元数据以及描述内容和工具来具体规范如何创造、存储、调整和分析数据档案，则数字科研数据没有丝毫价值[7]。

数据政策要求接受资助的科研人员以及论文作者，确保出版或者在项目中产生的数据具有可获取性[8]。这对科学家们以及科研机构如何持有科研数据产生了影响，而且对资助机构和期刊的相关政策建议和强制执行方式会有更大影响。举例来说，美国国家科学基金会(NSF)的《数据开放政策》要求项目申请者“在不增加成本以及过多时间的情况下，将原始数据、样本、物质材料和其他在 NSF 资助下的工作所创造或者产生的附加材料，与其他科研人员共享”[9]。在《数据管理计划》中 NSF 更进一步要求测量这项政策规定的执行情况[10]。德国科研基金会(DFG)自 2010 年开始的项目方案中具有类似的科研数据处理原则，据此，德国科研人员应该尽其所能地符合“现存标准和数据知识库”的要求[11]。类似条件也可在学术期刊出版商的“编辑政策”中发现，例如自然出版集团(NPG)要求“作者不设任何限制地提供材料、数据和有关协议给其他人”[12]。因此，科研数据的开放共享理应通过公共知识库来实现。

尽管科学家们认同为科学进步实施数据共享，有其共同潜在利益；然而多数人在实际执行的时候还是有所顾忌[13,14]。激励措施可以促进变化，例如数据的适当引用等[15]。因此，数据共享集成在学术交流中有很大的发展空间。

科研数据可以经过三种发布策略实施开放获取[16]:

(1) 科研数据独立发表：作为独立的信息对象，由知识库发布[17]。

(2) 科研数据与文本文献一同发表：此类情况又称数据论文[18]。

(3) 科研数据依附出版物发表(使出版物丰富化)：作为注释文本内容的材料，以丰富出版物内容的用途，作为一种说明文件[19]。

这些发布策略的共同特征是要求信息基础设施能够确保数据在最大程度上进行永久保存和开放共享。这种基础设施的名称包括：数据档案、数据中心、数字图书馆、数字典藏和其他名称等，我们称它们是科研数据知识库(RDR)。

目前为止，人们缺乏针对上述基础设施及其功能的全面概述。科研数据知识库的注册机制 re3data.org 改变了这种不利的局面。该项目已经在 2012 年开始对科研数据知识库进行索引，并且提供科研人员、资助机构、图书馆和出版商一份系统性概述异构 RDR 的全景观测。在 2013 年 7 月，re3data.org 的注册清单上已有 400 个科研数据知识库。其中的 288 个采用一份由 re3data.org 开发的特殊标识符号来详细描述自己。以下给出一个 RDR 的全景观测(见第 2 节)；此外，本文描述了注册机制的发展，在 re3data. org 上的功能，以及解释这套机制如何帮助人们确定适合科研数据存储与搜索的知识库(见第 3 节)。

2　科研数据知识库全景

欧盟理事会 2009 年的 ICT 报告总结“整个欧洲的数据知识库非常异质化，不过我们具有连贯的发展政策，作为克服这种零散状态的坚实基础，并且提供科研社群得以更好地管理、使用、存储和保存数据”[20]。理事会强调目前信息基础设施的全景观测恰当清晰地表达了对集成和同化科研数据服务的需求。

RDR 及其服务因为所存储内容的不同而各具特点，它们在不同情况下，为各式各样文件类型的获取与使用进行存储。然而对比科研数据的存储，人们更关注知识库提供科研出版物的标准规范。开放档案协

议(OAI)很早就建立了促进机构或者学科知识库的标准规范和网络，以提供文本信息对象的开放获取，如科研论文(前出版或者后出版)、学位论文等[21]。反之，RDR 社群缺乏可以比拟的标准规范。

直至今日，仅有少数研究调查了科研数据知识库的全球现况，例如 2010 年 Marcial & Hemminger 出版对 100 个 RDR 的调查研究[22]。而在 2011 年 Schaaf 有类似研究[23]。纵观各个学科已经出现大量不同的 RDR 项目，即便只看单一学科，生物医学也提供了为数可观的 RDR 数目，它们塑造了今日科研数据基础设施的全景。不仅如此，生物医学的数字基础设施也能够被其他专业学术社群获取。核酸研究期刊在 2013 年编辑的“分子生物数据库” (http://www.oxfordjournals.org/nar/database/a/)呈现 1 512 个生物科研数据存储的基础设施[24]。其中的 200 个基础设施通过在欧洲生命科学基础设施的范围内的生物信息项目(ELIXIR)。有来自 100 个研究机构至少 350 位员工在负责这 200 个知识库的运行。数以千计的科学家们组成的社群，使用了这类 RDR。这 200 个 RDR 每年直接成本大约是 3 000 万欧元[25]。为了生物医学 RDR 的可持续运作，ELIXIR 加入欧洲战略论坛科研基础设施(ESFRI)，有策略地推动和维持欧洲科研范围(European Research Area, ERA)作为国际科研中心的地位。从 2004 年启动时 ESFRI 就很清楚科研基础设施不仅只有物理基础设施，如科研舰队或粒子加速器，而且包括数字信息基础设施，像是“科研出版与数据库的电子档案系统”等[26]。

2.1 科研数据知识库的类型

以下是在分析 400 个 RDR 的基础上演化而来的 RDR 类型体系。基于机构和学科的学术文献知识库之间具有宽泛的差异性[27]，本文作者区分机构的、学科的、跨学科的，以及项目的 RDR[28]。以下说明 4 种知识库之间的类型差异。这种系统性梳理有助于全面鸟瞰科研数据基础设施在永久访问和重用上的不同概念和战略。

(1) 机构科研数据知识库

机构科研数据知识库由大学或者研究机构来运行。在大学层面上是跨学科规模。爱丁堡 DataShare (http://datashare.is.ed.ac.uk)是英国的一个机构 RDR 案例。爱丁堡大学在 DSpace 软件框架的基础上，制定“跨学科科研数据的网络数字知识库”[29]，并且在 2007 年到 2009 年完成开发[30]。截至 2013 年 3 月，知识库存储共 61 个数据集。慕尼黑大学开放数据(http://data.ub.uni-muenchen.de)是另外一个来自德国的机构 RDR 案例。自 2010 年起采用 ePrints 软件并且作为慕尼黑大学的所有成员的科研数据发布平台[31]。该知识库在 2013 年 3 月存储了 35 项数据集。

(2) 学科科研数据知识库

在学科 RDR 里比较突出的案例是 GenBank 和 PANGAEA。GenBank 的服务(http://www.ncbi.nlm.nih.gov/genbank)始于 1982 年[32]而且定位自己是“支持书目数据和生物注解的核苷酸序列的公共数据库”。美国国家生物技术信息中心(NCBI)运行这套基础设施，提供超过 25 万种的核苷酸序列信息[33]。

作为地球与环境科学的数据发布者(http://www.pangaea.de)，盘古大地(PANGAEA)的定位是“存档、发布、传播从地球系统研究得到的地理参照数据的开放获取图书馆”[34]。这个 RDR 是由 Alfred Wegener 极地与海洋研究所(AWI)和不来梅大学海洋环境科学中心(MARUM)来运行。PANGAEA 开始的时候是“古代气候数据中心”，它在 1994 年到 1997 年间受到德国联邦教育与研究部(BMBF)的资助[35]。在 2011 年 PANGAEA 存储“大约 50 万笔的地球科学领域的数据集”[36]。

(3) 跨学科科研数据知识库

除了机构和学科等方式，科研数据知识库还能服务跨学科需求。作为科研数据知识库案例之一的 Figshare(http://figshare.com)就“允许科研人员采取引用、搜索和共享的方式发布他们所有的数据”[37]。从 2011 年开始，Figshare 由 Macmillan 出版公司的数字科学部门来营运[38]。第二个案例 LabArchives (http://www.labarchives.com)是“基于网络的电子笔记本软件”，它由一家私人公司经营，允许科学家们“储存、组织和发布他们的科研数据”[39]。

(4) 项目科研数据知识库

特别的科研项目会产生特殊的科研数据，而它们导致各式各样的 RDR 状态。由 GFZ 德国地球科学研究中心运行的科学钻探数据库(SDDB)(http://www.scientificdrilling.org)可被称为典范。它提供产生自科学大陆钻探计划(ICDP)的钻探数据的开放重用[40]。

伯尼尔数字神殿项目(http://www.digitalpantheon.ch/repository)的 RDR 是另一个案例，罗马神殿的高解析度图像和可视化均可自由访问。

上述 4 种类型呈现了异构 RDR 的整体景观。它被实际运用在描述那些符合潜在用户需求的 RDR 服务。

2.2　科研数据管理服务与工具的需求

从科研人员的角度来看，目前存在各种各样影响学者们主动分享科研数据的障碍。目前 Kuipers 和 Van der Hoeven 的全面研究[13]、Tenopir 等[14]以及 ODE 项目[41]等的研究表明，数据共享的意愿与一个支持科研数据的基础设施密切相关。嵌入学术工作流程的知识库以及与之相关的激励措施，可以促进数据共享。Tenopir 等在一项超过 1 300 名科学家的调查中，得出结论："几乎所有学科的多数受访者……愿意把他们的科研数据至少放置一部分在一个没有限制的中央数据知识库内"。然而障碍之一是学者们缺乏对已有 RDR 的完整认识。

为此，re3data.org 付诸行动。今天，在大多数学科中，很难得知现有 RDR 的完整概况。尽管有像 OpenDOAR 开放获取知识库目录(http://www.opendoar.org)和 ROAR 开放获取知识库注册(http://roar.eprints.org)这样的机制，也只包含一小部分的科研数据知识库(小于 5%)的信息，因为这两个机制的重点是学术出版物的知识库。过去几年，像是 OAD 开放获取目录(http://oad.simmons. edu/oadwiki/Data_repo sitories)以及 DataCite(http:// www.datacite.org/repolist)等网站开始列举 RDR。然而，这些目录以及类似的服务只提供了关于 RDR 和它们服务的基础信息，像是简短描述那些维护中的知识库运行、学科和 URL 等。为了克服上述在用户调查[13,14,41]中所看到的障碍，有必要向科研人员、资助机构、图书馆以及出版商等提供 RDR 的系统性和易用性综述。这意味着，如果注册机制想要传递实实在在的信息，或者在科研数据重用的条件下实施开放共享，就有必要比现有的目录提供更为详细的 RDR 描述。科研人员想要知道怎么存储他们所拥有的数据，所以他们必需知道 RDR 特性的信息，像是：这个 RDR 上线多久了？它受到什么样的资助？这个 RDR 是否有其政策？谁负责这个 RDR 的运作？这些都是让科研人员得以信任 RDR 的必要信息。

3　注册科研数据知识库的 re3data.org 机制

科研人员即将高度关注 RDR 的存储与利用。然而，开放科学的技术与政策需求[42,43]，包括公共资助的科研数据与结果的开放获取等，如果缺乏可信、持久、可持续的基础设施来支持科研人员共享他们科研数据，则注定要失败。对 RDR 运行的调查显示，超过 5 年就不再为基础设施提供财务支持的安全期具有不确定性[13,44]。所以，当前的科研发展战略和长期资助计划还有许多需要积极填补的空缺。

欧盟委员会在 2010 年曾经委托的一项研究，描述在 2030 年的科研数据处理愿景。由此得知，科研人员需要能够"发现、获取和处理他们所需的数据"。此外，科研人员采集数据时，将会在国际标准的基础上进行"在可靠的知识库中存储他们可信赖的数据"[45]。另一方面，实现这个愿景的学术体系所面临的核心挑战是：因应数据共享增长的背景以及 RDR 的异构性。有鉴于此，科研数据知识库协议的注册机制 re3data.org(http://re3data.org)针对 RDR 目录进行开发与运行。这项计划旨在对所有领域的 RDR 基于注册机制进行索引化和结构化描述。用信息图标来描述每个 RDR 的基本特征，这样就能创造快速和容易使用系统的附加价值。

德国地球科学研究中心(GFZ)的图书馆与信息服务部门(LIS)、柏林洪堡大学的图书馆与信息科学学系，以及卡尔斯鲁厄理工学院(KIT)图书馆是 re3data.org 项目的核心成员。这三个项目伙伴与德国网络信息协议组织(DINI)具有长期固定合作关系。在 DINI 资助下的科研数据政策报告于 2009 年发布[46]。德国科研基金会(DFG)资助了自 2012 年 1 月到 2013 年 12 月的第一期项目。

该项目的主要目标是提供数据供应者和数据使用者双方科研人员对 RDR 异构全景的指引，并且服务科研资助者和基础设施维护者，如数据中心和学术图书馆。进一步来说，re3data.org 旨在建立更为集中和集成的"数据知识库的生态系统"[47]；注册机制描绘世界范围内的 RDR 发展状况。这种全球概览也可用于协助那些 RDR 尚未发达的学科领域。

起先 re3data.org 列表上只有少数 RDR 以及基本信息，像是知识库名称、维护者和学科专业等。在

2012年12月项目收集和记录了将近400个存储科研数据的基础设施。目前的开放获取目录(OAD)列表就采用了这套全景观测结果。

re3data.org 项目的三个合作伙伴各自独立检测随机选取20次的RDR。第一次的分析确认了一幅极端异质的RDR全景以及成为创作RDR描述框架草稿的基础。因为缺乏一个合适框架，所以促使re3data.org发展一个新的描述RDR的元数据框架。在第二阶段，这个架构对接类似的元数据框架，调整了图标元素，并且介绍RDR的基本条件。

第一版描述RDR的关键术语在2012年7月的一份文件中发表[48]。对图标的评论取自电子邮件反馈以及项目网站，在确保符号体系发展的透明性的基础上，它获得了RDR社群的参与和接纳。

各项反馈十分正面，而且在一些案例中有非常缜密的阐述。项目得到来自 reBIND (http://rebind.bgbm.org)、DataCite(http://www.datacite.org)和OpenAIR Eplus (http://www.openaire.eu)以及其他反馈。项目小组分析和讨论了所有意见并且建议纳入关键术语的考量，以指导修改在2012年12月所公布的核心要素第二版[49]。这些图标要素(符号体系)包括以下几个方面(见图1):

(1) 一般信息(如RDR的简短描述、内容类型、关键词汇等);

(2) 知识库群(如受资助的机构责任、内容或者技术议题等);

(3) 政策(如RDR政策，包括它们的URL等);

(4) 法律议题(数据库以及数据集的许可证等);

(5) 技术标准(如应用程序界面、数据集版本、RDR软件等);

(6) 质量标准和服务(如证书、审计过程等)。

由于不同学术社群的异构性需求以及RDR标准的共同匮乏，认证条件与RDR审计程序受到检验[50-54]，其中若干条件存在不普遍适用RDR的情况。所以，需要适当降低RDR加入re3data.org注册的门槛。然而，为了让知识库在re3data.org能被索引，科研数据的获取和许可的细节缺一不可。如果您的RDR满足了这些基本要求，则可以被索引和被评论。



图1 科研数据知识库的关键要素以及re3data.org的图标

图标集凸显了知识库的主要特性，如图2所示。这套图标系统帮助用户去选择适合存储他们数据的知识库。科研人员在re3data.org中可以清楚看到每个RDR的访问与使用条件以及其他特性。

该网站(http://www.re3data.org/faq)解释了图标和它们的含义。这些图标引起RDR运行机构在re3data.org上注册的意愿。不过，符号系统不只是对科研人员有用，而且也对RDR维护者有用，有助于后者比较知识库的优缺点。这也使得re3data.org成为一项随着RDRs而不断更新和改版的实用工具。

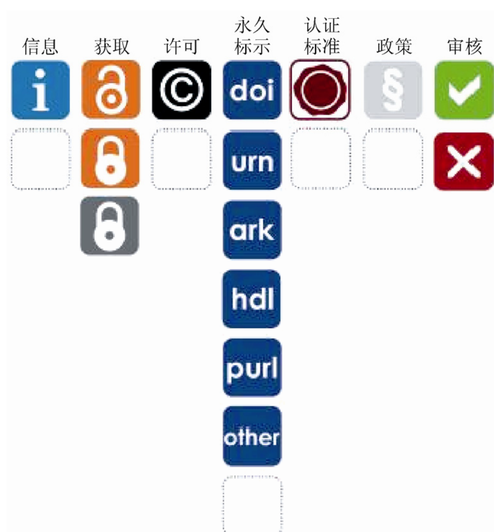


图 2　描绘科研数据权益的符号系统

简单质朴的 re3data.org 搜索设计(见图 3)会让每次搜索结果包括: RDR 的名称、知识库覆盖的主题、一些描述基础设施的图标(见图 4), 以及是否已经被 re3data.org 审核通过等信息。

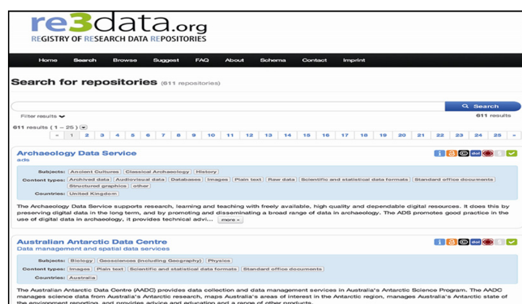


图 3　re3data.org 的搜索主页

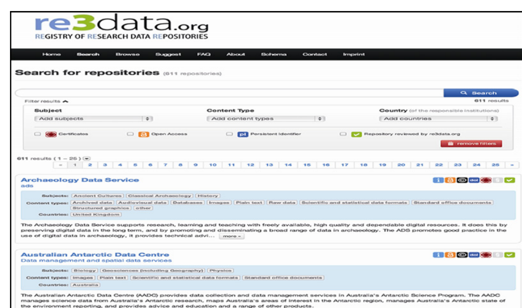


图 4　搜索后的点击列表

信息过滤是检索结果的凝练, 它采用: 主题、内容类型、国家、认证、开放获取、永久标识符以及审核情况等次级选项, 来产生搜索分页。通过点击知识库名称, 可以浏览各个 RDR 入口网站如图 5 所示:

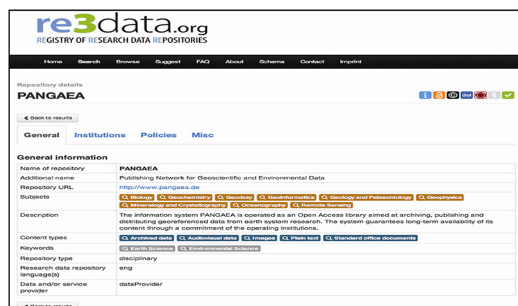


图 5　科研数据知识库的详细描述

RDR 维护者可以经过一道简单的申请方式, 建议 re3data.org 收录他们的数字基础设施。项目小组会列举和检查在目录上提交的知识库。当满足最低收录条件时, 知识库就被索引, 这意味着它们可以提供数据获取以及在网页上有明确的术语解释。我们在实践过程中考虑到: 如果过分区隔 RDR 网站的结构, 会耗费索引过程的时间; 而且目前仅有少数 RDR 具有服务政策、指定社群和使用条款, 还有某些 RDR 要求联系维护者才能拿到这些信息。鉴于这类情况的频繁发生, 我们优化了 re3data.org 工作流程, 并且改善了 RDR 维护者的反馈渠道。

4　展　望

随着负责数字议程(http://ec.europa.eu/digital-agenda)的欧盟委员 Kroes 疾呼“我们要开启开放科学的时代”宣言, 凸显了开放性(Openness)就是数字科学的范式[42]。这要求发展一种永久性的信息基础设施, 好让科学家们能够共享他们的科研数据, 并且让下一代科研人员能够继续获取与重用科研数据。

所有 re3data.org 项目合作伙伴资助了注册机制的长期运行。基于利益相关者的反馈, re3data.org 将继续发展科研数据管理的新功能与服务。对此安排与 DataCite 在 2012 年春季签署了合作备忘录。作为科研数据的永久唯一标示符倡议者之一的 DataCite 是德国科研基金会资助的数据出版项目成果, 也是 re3data.org 的联盟成员之一[17]。在这种合作模式下, 两个团体之间的信息交流显得极为重要。目前正在探讨像是 Databib(http://databib.org)等的相关协议。为了促进注册机制的技术和结构化发展, re3data.org 和它的项目伙伴将继续促进 RDR 更紧密的集成和更大规

模的连贯性。

尽管 re3data.org 尚在起步阶段，截至 2013 年 7 月已有 400 个 RDR 编入 re3data.org 的索引，超过 288 个被审核通过。项目下一阶段的工作集中在改善可用性和实施新功能。超越登记注册的发展，本项目追求科研数据知识库的标准和网络。本项目努力促使在创作共用许可 CC0(许可授权：公共知识共享)下，所有注册的元数据得以开放使用。与此同时，re3data.org 为开放科学开辟了一条实践之道。

(注：原文作者致谢开发 re3data.org 期间参与项目的成员：Gabriele Kloska, Evelyn Reuter, Jessika Rücknagel, Markus Schnalke, Edeltraud Schnepf, Angelika Semrau, Shaked Spier。)

参考文献：

[1] Berlin Declaration. Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities [OL]. [2009-10-27]. http://oa. mpg. de/openaccess-berlin/ berlindeclaration. html.

[2] Organisation for Economic Co-operation and Development (OECD) . Principles and Guidelines for Access to Research Data from Public Funding [OL]. [2013-05-16]. http://www. oecd. org/dataoecd/9/61/38500813. pdf.

[3] The Royal Society. Science as an Open Enterprise [OL]. [2013-05-16]. http://royalsociety. org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE. pdf.

[4] European Commission. Commission Recommendation on Access to and Preservation of Scientific Information [OL]. [2013-05-16]. http://ec. europa. eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en. pdf.

[5] Office of Science and Technology Policy. Increasing Access to the Results of Federally Funded Scientific Research [OL]. [2013-05-16]. http://www. whitehouse. gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013. pdf.

[6] European Commission. Communication from the Commission to the European Parliament, the European Economic and Social Committee and the Committee of the Regions [OL]. [2013-05-16]. http://ec. europa. eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en. pdf.

[7] Kindling M, Schirmbacher P. “Die Digitale Forschungswelt” als Gegenstand der Forschung [J]. Information - Wissenschaft & Praxis, 2013, 64: 127-136.

[8] Pampel H, Bertelmann R. “Data Policies” im Spannungsfeld Zwischen Empfehlung und Verpflichtung [A]. //Büttner S, Hobohm H C, Müller L, (edt). Handbuch Forschungsdatenmanagement [M]. Bad Honnef: Bock+Herchen, 2011:49-61.

[9] National Science Foundation(NSF). Proposal and Award Policies and Procedures Guide. Chapter VI - Other Post Award Requirements and Considerations [OL]. [2013-05-16]. http://www. nsf. gov/pubs/policydocs/ pappguide/nsf11001/aag_6. jsp#VID4.

[10] National Science Foundation(NSF). Proposal and Award Policies and Procedures Guide. Grant Proposal Guide. Chapter II - Proposal Preparation Instructions [OL]. [2013-05-16]. http://www. nsf. gov/pubs/policydocs/pappguide/nsf11001/gpg_2. jsp#dmp.

[11] Deutsche Forschungsgemeinschaft. Proposal Preparation Instructions. DFG form 54. 01. [OL]. [2013-05-16]. http://www. dfg. de/formulare/54_01/54_01_en. pdf.

[12] Nature. Availability of Data and Materials [OL]. [2013-05-16]. http://www.nature.com/authors/policies/availability. html.

[13] Kuipers T, Van der Hoeven J. Insight into Digital Preservation of Research Output in Europe [R/OL]. [2013-05-16]. http://www. parse-insight. eu/downloads/PARSE-Insight_D3-4_SurveyReport_final_hq. pdf.

[14] Tenopir C, Allard S, Douglass K, et al. Data Sharing by Scientists: Practices and Perceptions [J]. PLoS ONE, 2011, 6(6): e21101.

[15] Nature Biotechnology. Credit Where Credit is Overdue [J]. Nature Biotechnology, 2009, 27: 579.

[16] Pampel H, Dallmeier-Tiessen S. Open Research Data - From Vision to Practice [A]. //Bartling S, Friesike S, (edt). Opening Science–The Evolving Guide on How the Internet Is Changing Research, Collaboration and Scholarly Publishing [M]. Heidelberg: Springer, 2009.

[17] Klump J, Bertelmann R, Brase J, et al. Data Publication in the Open Access Initiative [J]. Data Science Journal, 2006, 5: 79-83.

[18] Chavan V, Penev L. The Data Paper: A Mechanism to Incentivize Data Publishing in Biodiversity Science [J]. BMC Bioinformatics, 2011, 12: S2. DOI:10. 1186/1471-2105-12-S15-S2.

[19] Woutersen-Windhouwer S, Brandsma R, Hogenaar A, et al. Enhanced Publications?: Linking Publications and Research Data in Digital Repositories [OL]. [2013-05-16]. http://dare. uva. nl/aup/nl/record/316849.

[20] European Commission. ICT Infrastructures for e-Science

[OL]. [2013-05-16]. http://eur-lex. europa. eu/LexUriServ/LexUriServ. do?uri=COM:2009:0108:FIN:EN:PDF.

[21] Lagoze C, Sompel H Van De. The Open Archives Initiative: Building a Low-Barrier Interoperability Framework [OL]. [2013-05-16]. http://public.lanl.gov/herbertv/papers/Papers/2001/JCDLlagoze. pdf.

[22] Marcial L H, Hemminger B M. Scientific Data Repositories on the Web: An Initial Survey [J]. Journal of the American Society for Information Science and Technology, 2010, 61: 2029-2048.

[23] Schaaf I. Forschungsdaten im Netz – Untersuchung Von Online Verfügbaren Repositorien. Hochschule der Medien Stuttgart [A]. //Schäfer L, Pampel H, Klump J, et al. Bericht Symposium"Forschungsdaten-Infrastrukturen (FDI 2013)" [C]. 2013. DOI:10. 2312/radieschen_003.

[24] Fernández-Suárez X M, Galperin M Y. The 2013 Nucleic Acids Research Database Issue and the Online Molecular Biology Database Collection [OL]. [2013-05-16]. http://www. ncbi. nlm. nih. gov/pubmed/23203983.

[25] ELIXIR . The ELIXIR Strategy for Data Resources [OL]. [2013-05-16]. http://www. elixir-europe. org/prep/sites/elixir-europe.org.prep/files/documents/reports/elixir_strategy_for_data_resources_report. pdf.

[26] European Strategy Forum on Research Infrastructures. European Roadmap for Research Infrastructures. Report 2006 [OL]. [2013-05-16]. http://ec.europa.eu/research/infrastructures/pdf/esfri/esfri_roadmap/roadmap_2006/esfri_roadmap_2006_en. pdf.

[27] Suber P. Open Access. Cambridge, Massachusetts [OL]. [2013-03-15]. The MIT Press. http://mitpress.mit.edu/books/open-access.

[28] Pampel H, Goebelbecker H J, Vierkant P. re3data. org: Aufbau eines Verzeichnisses von Forschungsdaten-Repositorien. Ein Werkstattbericht [A]. //Mittermaier B, Wissen V. Daten, Menschen, Systeme. WissKom 2012. Jülich: Verlag des Forschungszentrums Jülich, 2012: 61-73.

[29] Edinburgh DataShare. What is Edinburgh DataShare? [OL]. [2013-05-16]. http://datashare.is.ed.ac.uk.

[30] Rice R. DISC-UK DataShare [OL]. [2013-05-16]. http://ie-repository.jisc.ac.uk/336/.

[31] Schallehn V. Open Data LMU. Universitätsbibliothek München [OL]. [2013-05-16]. http://www. ub. uni-muenchen. de/no_cache/aktuelles/einzelne-nachricht/article/open-data-lmu/.

[32] Cravedi K. GenBank Celebrates 25 Years of Service with Two-Day Conference [OL]. [2013-05-16]. http://www.nih. gov/news/health/apr2008/nlm-03.htm.

[33] Benson D A, Karsch-Mizrachi I, Clark K, et al. GenBank [J]. Nucleic Acids Research, 2012, 40: D48-D53.

[34] PANGAEA. About/Imprint [OL]. [2013-05-16]. http://www. pangaea.de/about/.

[35] Diepenbroek M, Grobe H, Reinke M, et al. Data Management of Proxy Parameters with PANGAEA [A]. //Fischer G, Wefer G. Use of Proxies in Paleoceanography [M]. New York: Springer-Verlag, 1999:715-727.

[36] Schindler U, Diepenbroek M, Grobe H. PANGAEA-Research Data Enters Scholarly Communication [J]. Geophysical Research Abstracts, 2012, 14: EGU2012–13378-1.

[37] Figshare. FAQs [OL]. [2013-05-16]. http://figshare.com/faqs.

[38] Fenner M. Figshare: Interview with Mark Hahnel [OL]. [2013-05-16]. http://blogs. plos. org/mfenner/2012/02/16/figshare-interview-with-mark-hahnel/.

[39] LabArchives. Electronic Notebook Software for Laboratory Notebooks FAQ [OL]. [2013-05-16]. http://www.labarchives.com/faqs.php.

[40] Klump J, Conze R. The Scientific Drilling Database (SDDB) -Data from Deep Earth Monitoring and Sounding [OL]. [2013-05-16]. http://www.iodp.org/images/stories/downloads/sd4_07.pdf#page=32.

[41] Dallmeier-Tiessen S, Darby R, Gitmans K, et al. Summary of the Studies, Thematic Publications and Recommendations [OL]. [2013-05-16]. http://www. alliancepermanentaccess. org/wp-content/plugins/download-monitor/download.php?id=Summaryofthestudies%2Cthematicpublicationsandrecommendations.

[42] Kroes N. Opening Science Through e-Infrastructures [OL]. [2013-05-16]. http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/258.

[43] Kroes N. Making Open Access a Reality for Science [OL]. [2013-05-16]. http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/392.

[44] Pfeiffenberger H, Pampel H, Schäfer A, et al. Report and Strategy on Annotation, Reputation and Data Quality [OL]. [2013-05-16]. http://www.alliancepermanentaccess. org/wp-content/plugins/download-monitor/download.php?id=D26.1ReportandStrategyonAnnotation%2CReputationandDataQuality.

[45] High Level Expert Group on Scientific Data. Riding the Wave. How Europe can Gain from the Rising Tide of Scientific Data [OL]. [2013-05-16]. http://cordis. europa. eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report. pdf.

[46] Dallmeier-Tiessen S, Dobratz S, Gradmann S, et al. Positionspapier Forschungsdaten. 1. 0 ed [OL]. [2013-05-16]. http://nbn-resolving. de/urn:nbn:de:kobv:11-10098082.
[47] Graaf M, Waaijers L. A Surfboard for Riding the Wave-Towards a Four Country Action Programme on Research Data [OL]. [2013-05-16]. http://www. knowledge-exchange. info/ Default. aspx?ID=469.
[48] Vierkant P, Spier S, Rücknagel J, et al. Vocabulary for the Registration and Description of Research Data Repositories [OL]. [2013-05-16]. http://dx. doi. org/10. 2312/re3. 001.
[49] Vierkant P, Spier S, Rücknagel J, et al. Vocabulary for the Registration and Description of Research Data Repositories (Version 2.0) [OL]. [2013-05-16]. http://dx. doi. org/10. 2312/re3. 002.
[50] Braun K, Buddenbohm S, Dobratz S, et al. DINI Certificate Document and Publication Services 2010 [OL]. [2013-05-16]. http://nbn-resolving. de/urn:nbn:de:kobv:11-100182800.
[51] Consultative Committee for Space Data Systems (CCSDS). Audit and Certification of Trustworthy Digital Repositories [OL].[2013-05-16].http://public.ccsds.org/publications/archive/652x0m1.pdf.
[52] Data Seal of Approval. Data Seal of Approval. Guidelines Version 1 [OL]. [2013-05-16]. http://assessment.datasealofapproval. org/documentation/.
[53] ESF & EUROHORCs (2011) Basic Requirements for Research Infrastructures in Europe [OL]. [2013-05-16]. http://www.dfg.de/download/pdf/foerderung/programme/wgi/basic_requirements_research_infrastructures. pdf.
[54] ICSU World Data System (2011) Certification of World Data System Members Introduction [OL]. [2013-05-16]. http://icsu-wds.org/images/files/Certification_summary_6_Jul_2011.pdf.

(通讯作者：顾立平 E-mail：gulp@mail.las.ac.cn)

Making Research Data Repositories Visible: The re3data.org Registry

Heinz Pampel[1] Paul Vierkant[2] Frank Scholze[3] Roland Bertelmann[1] Maxi Kindling[2] Jens Klump[1]
Hans-Jürgen Goebelbecker[3] Jens Gundlach[3] Peter Schirmbacher[2] Uwe Dierolf[3]
[1](Deutsches GeoForschungsZentrum GFZ, Library and Information Services (LIS), Potsdam 14473, Germany)
[2](Humboldt-Universität zu Berlin, Berlin School of Library and Information Science, Berlin 10099, Germany)
[3](Karlsruhe Institute of Technology (KIT), KIT Library, Karlsruhe 76131, Germany)

Abstract: Researchers require infrastructures that ensure a maximum of accessibility, stability and reliability to facilitate working with and sharing of research data. Such infrastructures are being increasingly summarized under the term Research Data Repositories (RDR). The project re3data.org–Registry of Research Data Repositories–has begun to index research data repositories in 2012 and offers researchers, funding organizations, libraries and publishers an overview of the heterogeneous research data repository landscape. In July 2013 re3data.org lists 400 research data repositories and counting. 288 of these are described in detail using the re3data.org vocabulary. Information icons help researchers to easily identify an adequate repository for the storage and reuse of their data. This article describes the heterogeneous RDR landscape and presents a typology of institutional, disciplinary, multidisciplinary and project-specific RDR. Further the article outlines the features of re3data.org, and shows how this registry helps to identify appropriate repositories for storage and search of research data.

Keywords: Research data Research data management Open access Institutional repositories development Open science Research data repositories Information management